

09/829,831

**REMARKS**

In view of the following discussion, the Applicants submit that none of the claims now pending in the application is obvious under the provisions of 35 U.S.C. § 103. Thus, the Applicants believe that all of these claims are now in allowable form.

**I. REJECTION OF CLAIMS 1, 2, 11, 12 AND 21 UNDER 35 U.S.C. § 103**

Claims 1, 2, 11, 12 and 21 stand rejected as being obvious over the Lee patent (U.S. 5,617,507, hereinafter "Lee") in view of the Silverman patent (U.S. 5,890,117, hereinafter "Silverman"). The Applicants respectfully traverse the rejection.

Lee teaches a text-to-speech synthesis system that is adapted for receiving text input and generating corresponding synthesized speech output. Specifically, the system receives text input, e.g., from a computer keyboard, and analyzes the syntax of the text in order to convert the text signal to a string of phonetic transcriptive symbols. The system then generates intonation pattern data and stress pattern data so that the appropriate prosody (e.g., intonation and stress) can be applied to the string of phonetic transcriptive symbols. The string of phonetic transcriptive symbols, including the applied prosody, is then output to a speech segment concatenation subsystem, which generates audible synthetic speech output.

Silverman also teaches a text-to-speech synthesis system that produces synthesized speech from input text signals. Text input, e.g., received from a touch-tone telephone keypad, is processed by a text processor. The text is then embedded with prosodic indicia or markers that specify, to a speech synthesizer, the desired prosody for the input text. The synthesizer then "speaks" the input text, applying the appropriate prosody to the synthetic speech.

The Examiner's attention is directed to the fact that Lee and Silverman, singularly or in combination, fail to disclose or suggest the novel invention of extracting prosodic features from an input speech signal, as claimed in Applicants' independent claims 1, 11 and 21. Specifically, Applicants' claims 1, 11 and 21 positively recite:

09/829,831

1. A method for processing a speech signal comprising:  
extracting prosodic features from a speech signal;  
modeling the prosodic features to identify at least one speech endpoint; and  
producing an endpoint signal corresponding to the occurrence of the at least one speech endpoint. (Emphasis added)
11. Apparatus for processing a speech signal comprising:  
a prosodic feature extractor for extracting prosodic features from the speech signal;  
a prosodic feature analyzer for modeling the prosodic features to identify at least one speech endpoint; and  
an endpoint signal producer that produces an endpoint signal corresponding to the occurrence of the at least one speech endpoint. (Emphasis added)
21. An electronic storage medium for storing a program that, when executed by a processor, causes a system to perform a method for processing a speech signal comprising:  
extracting prosodic features from a speech signal;  
modeling the prosodic features to identify at least one speech endpoint; and  
producing an endpoint signal corresponding to the occurrence of the at least one speech endpoint. (Emphasis Added)

In one embodiment, the Applicants' invention is directed to method for applying prosody-based endpointing to a speech signal. Conventional speech processing techniques that are used to provide signals, based on spoken words or commands (e.g., for controlling devices or software programs), typically are characterized by an inability or difficulty in locating suitable speech segments within the spoken input for processing. Typical endpointing techniques identify the completion of a speech segment or utterance by measuring pauses in the given speech signal. However, since spoken language is not typically produced with such explicit indicators, typical endpointing techniques may misinterpret normal fluctuations in the rhythm of speech, such as mid-sentence pauses, to indicate the completion of an utterance. The resultant translation of a spoken command may therefore be fraught with inaccuracies.

The Applicants' invention facilitates the translation of spoken input by extracting prosodic features from an input speech signal. The prosodic features are then modeled

09/829,831

In order to identify at least one endpoint in the input speech signal. The extraction of prosodic features yields more reliable endpoint identification results than conventional endpointing techniques (e.g., relying on measured pauses), because it enables the speech processing process to account for and consider natural speech characteristics such as changes in rhythm and pitch.

In contrast, both Lee and Silverman teach generating and embedding prosodic indicia into a text string in order to produce synthesized speech having a rhythm or intonation that more closely resembles natural speech. Thus, Lee and Silverman, singularly or in combination, fail to obviate Applicants' invention.

Specifically, both Lee and Silverman teach methods that start with text input and process the input to produce synthetic speech output. To this end, both Lee and Silverman only teach generating or embedding prosodic features into a received text string, e.g., in order to produce more natural sounding synthetic speech output. Neither Lee nor Silverman addresses the need to extract prosodic features from speech input in order to identify endpoints in the input speech for speech processing. In fact, both Lee and Silverman teach away from the Applicants' claimed invention, as they teach generating and inserting prosody for artificial speech and not extracting prosody from real speech. Lee and Silverman thus fail, singularly and in combination, to teach or make obvious a method for processing an input speech signal wherein prosodic features are extracted from the speech signal and modeled to identify at least one speech endpoint, as positively claimed by the Applicants in claims 1, 11 and 21. Therefore, the Applicants submit that independent claims 1, 11 and 21 fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

Dependent claims 2 and 12 depend respectively from claims 1 and 11, and recite additional features therefore. As such, and for the exact same reason set forth above, the Applicants submit that claims 2 and 12 are not made obvious by the teachings of Lee in view of Silverman. Therefore, the Applicants submit that dependent claims 2 and 12 also fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

09/829,831

**II. REJECTION OF CLAIMS 3, 4, 13 AND 14 UNDER 35 U.S.C. § 103**

Claims 3, 4, 13 and 14 stand rejected as being obvious over Lee in view of Silverman and further in view of the Chihara patent (U.S. 6,470,316, hereinafter "Chihara I"). The Applicants respectfully traverse the rejection.

Lee and Silverman have been discussed above. Chihara I, like Lee and Silverman, teaches a text-to-speech synthesis system that is adapted for receiving text input including Kanji and/or Kana characters (Chinese characters and Japanese syllabary) and generating corresponding synthesized speech output. Specifically, the system receives text input and refers to a word dictionary to determine the appropriate reading, accents and intonation of the input text. The system then generates a string of phonetic and prosodic symbols. A prosody generation module sets a plurality of parameters (e.g., pitch frequency pattern, phoneme duration, etc.) for subsequently generated synthesized speech, thereby producing more natural sounding synthesized speech. In one embodiment, one of the parameters that the prosody generation module sets is the length of pauses to be inserted into the synthesized speech.

The Examiner's attention is directed to the fact that Chihara I, singularly or in combination with Lee and Silverman, fails to disclose or suggest the novel invention of extracting prosodic features from an input speech signal, as claimed in Applicants' independent claims 1 and 11, from which claims 3, 4, 13 and 14 depend. Applicants' claims 1 and 11 have been recited above.

As discussed above with reference to the rejection under Lee in view of Silverman, the Applicants' invention facilitates the translation of spoken input by extracting prosodic features from an input speech signal. The prosodic features are then modeled in order to identify at least one endpoint in the input speech signal.

In contrast, Chihara I, like both Lee and Silverman, teaches generating and embedding prosodic indicia into a received text string in order to produce synthesized speech having a rhythm or intonation that more closely resembles natural speech. Thus, Lee, Silverman and Chihara I, singularly or in combination, fail to obviate Applicants' invention.

09/829,831

Specifically, Chihara I, like Lee and Silverman, teaches a method that starts with text input and processes the input to produce synthetic speech output. Thus, Chihara I only teaches generating or embedding prosodic features into a text string, not extracting prosodic features from a speech signal, e.g., in order to identify endpoints in the input speech for speech processing. Lee, Silverman and Chihara I thus fail, singularly and in combination, to teach or make obvious a method for processing an input speech signal wherein prosodic features are extracted from the speech signal and modeled to identify at least one speech endpoint, as positively claimed by the Applicant in claims 1 and 11. Therefore, the Applicants submit that Independent claims 1 and 11 fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

Dependent claims 3, 4, 13 and 14 depend from claims 1 and 11, and recite additional features therefore. As such, and for the exact same reason set forth above, the Applicants submit that claims 3, 4, 13 and 14 are not made obvious by the teachings of Lee in view of Silverman and further in view of Chihara. Therefore, the Applicants submit that dependent claims 3, 4, 13 and 14 also fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

### **III. REJECTION OF CLAIMS 5 AND 15 UNDER 35 U.S.C. § 103**

Claims 5 and 15 stand rejected as being obvious over Lee in view of Silverman and Chihara I and further in view of the Lin patent (U.S. 4,799,261, hereinafter "Lin"). The Applicants respectfully traverse the rejection.

Lee, Silverman and Chihara I have been discussed above. Lin teaches a speech synthesis system that is adapted for receiving text or speech input and generating corresponding synthesized speech output. Generally, the system receives text or speech input and encodes this data with phonological linguistics indicia (e.g., allophone indicia, syllable pitch pattern indicia and syllable duration pattern indicia) in order to produce synthesized speech output having a more natural quality. In one embodiment, syllable pitch data for the synthesized output is generated by comparing the input data against a plurality of stored pitch patterns.

09/829,831

The Examiner's attention is directed to the fact that Lin, singularly or in combination with Lee, Silverman and Chihara I, fails to disclose or suggest the novel invention of extracting prosodic features from an input speech signal and modeling the prosodic features to identify one or more speech endpoints, as claimed in Applicants' independent claims 1 and 11, from which claims 5 and 15 respectively depend. Applicants' claims 1 and 11 have been recited above.

As discussed above, the Applicants' invention facilitates the translation of spoken input by extracting prosodic features from an input speech signal. The prosodic features are then modeled in order to identify at least one endpoint in the input speech signal. This results, for example, in more accurate speech-to-text translation.

In contrast, Lin teaches generating and embedding prosodic indicia into a received text or speech string in order to produce synthesized speech having a rhythm or intonation that more closely resembles natural speech. Thus, Lee, Silverman, Chihara I and Lin, singularly or in combination, fail to obviate Applicants' invention.

Specifically, Lin, like Lee, Silverman and Chihara I, teaches a method that processes input data to produce synthetic speech output. To this end, Lin teaches generating or embedding prosodic features into the input data to produce synthesized output, not extracting prosodic features from a speech signal, e.g., in order to identify endpoints in the input speech for speech processing. Lee, Silverman, Chihara I and Lin thus fail, singularly and in combination, to teach or make obvious a method for processing an input speech signal wherein prosodic features are extracted from the speech signal and modeled to identify at least one speech endpoint, as positively claimed by the Applicant in claims 1 and 11. Therefore, the Applicants submit that independent claims 1 and 11 fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

Dependent claims 5 and 15 depend respectively from claims 1 and 11, and recite additional features therefore. As such, and for the exact same reason set forth above, the Applicants submit that claims 5 and 15 are not made obvious by the teachings of Lee in view of Silverman and Chihara and further in view of Lin. Therefore, the

09/829,831

Applicants submit that dependent claims 5 and 15 also fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

#### **IV. REJECTION OF CLAIMS 6 AND 16 UNDER 35 U.S.C. § 103**

Claims 6 and 16 stand rejected as being obvious over Lee in view of Silverman and Chihara I and further in view of a second Chihara patent (U.S. 6,625,575, hereinafter "Chihara II"). The Applicants respectfully traverse the rejection.

Lee, Silverman and Chihara I have been discussed above. Chihara II, like Chihara I, teaches a text-to-speech system that is adapted for receiving text input and generating corresponding synthesized speech output. Generally, the system receives text input generates corresponding prosodic indicia, including pitch and intonation patterns, which are embedded into the text string. The system generates synthetic speech output based on the text and generated prosodic indicia.

The Examiner's attention is directed to the fact that Chihara II, singularly or in combination with Lee, Silverman and Chihara I, fails to disclose or suggest the novel invention of extracting prosodic features from an input speech signal and modeling the prosodic features to identify one or more speech endpoints, as claimed in Applicants' independent claims 1 and 11, from which claims 6 and 16 respectively depend. Applicants' claims 1 and 11 have been recited above.

As discussed above, the Applicants' invention facilitates the translation of spoken input by extracting prosodic features from an input speech signal. The prosodic features are then modeled in order to identify at least one endpoint in the input speech signal.

In contrast, Chihara II teaches generating and embedding prosodic indicia into a received text string in order to produce synthesized speech having a rhythm or intonation that more closely resembles natural speech. Thus, Lee, Silverman, Chihara I and Chihara II, singularly or in combination, fail to obviate Applicants' invention.

Specifically, Chihara II, like Lee, Silverman and Chihara I, teaches a method that processes input text to produce synthetic speech output. To this end, Chihara II teaches generating or embedding prosodic features into the input data to produce

09/829,831

synthesized output, not extracting prosodic features from a speech signal, e.g., in order to identify endpoints in the input speech for speech processing. Lee, Silverman, Chihara I and Chihara II thus fail, singularly and in combination, to teach or make obvious a method for processing an input speech signal wherein prosodic features are extracted from the speech signal and modeled to identify at least one speech endpoint, as positively claimed by the Applicant in claims 1 and 11. Therefore, the Applicants submit that independent claims 1 and 11 fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

Dependent claims 6 and 16 depend respectively from claims 1 and 11, and recite additional features therefore. As such, and for the exact same reason set forth above, the Applicants submit that claims 6 and 16 are not made obvious by the teachings of Lee in view of Silverman and Chihara and further in view of Chihara II. Therefore, the Applicants submit that dependent claims 6 and 16 also fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

#### **V. REJECTION OF CLAIMS 7-10 AND 17-20 UNDER 35 U.S.C. § 103**

Claims 7-10 and 17-20 stand rejected as being obvious over Lee in view of Silverman and further in view of the Neumeyer patent (U.S. 6,226,611, hereinafter "Neumeyer"). The Applicants respectfully traverse the rejection.

As Neumeyer was filed on January 26, 2000 and issued on May 1, 2001, after Applicants' filing date of April 10, 2001, Neumeyer is a §102(e) type reference. Neumeyer is a continuation of U.S. Patent No. 6,055,498. Both the '498 patent and the Applicants' invention were commonly assigned to SRI International (reel/frame 9474/0501 and reel/frame 012018/0894, respectively) at the time Applicants' invention was made; thus, Neumeyer does not preclude patentability of the present invention under the provisions of 35 U.S.C. §103(c). MPEP 706.02(I)(1).

As discussed, Lee and Silverman do not teach, show or suggest the Applicants' claimed invention. Moreover, as Neumeyer cannot be properly combined with Lee and Silverman to obviate the Applicants' claims, the Applicants submit that independent claims 1 and 11 fully satisfy the requirements of 35 U.S.C. §103 and are patentable



09/829,831

thereunder. Moreover, dependent claims 7-10 and 17-20 depend respectively from claims 1 and 11, and recite additional features therefore. As such, and for the exact same reason set forth above, the Applicants submit that claims 7-10 and 17-20 are not made obvious by the teachings of Lee in view of Silverman and further in view of Neumeyer. Therefore, the Applicants submit that dependent claims 7-10 and 17-20 also fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

#### **VI. CLAIM AMENDMENTS**

Claims 19 and 20 have been voluntarily amended in order to correct minor typographical errors. Specifically, both claim 19 and claim 20 have been amended to recite an "apparatus," replacing a "method", in order to more clearly reflect the claims' dependence on independent claim 11.

#### **Conclusion**


Thus, the Applicants submit that all of these claims now fully satisfy the requirements of 35 U.S.C. §103. Consequently, the Applicants believe that all of these claims are presently in condition for allowance. Accordingly, both reconsideration of this application and its swift passage to issue are earnestly solicited.

If, however, the Examiner believes that there are any unresolved issues requiring the issuance of a final action in any of the claims now pending in the application, it is requested that the Examiner telephone Mr. Kin-Wah Tong, Esq. at (732) 530-9404 so that appropriate arrangements can be made for resolving such issues as expeditiously as possible.

Respectfully submitted,

7/2/04  
Date

Moser, Patterson & Sheridan, LLP  
595 Shrewsbury Avenue  
Shrewsbury, New Jersey 07702

  
Kin-Wah Tong, Reg. No. 39,400  
(732) 530- 9404